Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-023-04555-3

ORIGINAL RESEARCH

Hybrid model for early identification post-Covid-19 sequelae

Evandro Carvalho de Andrade[1,2] · Luana Ibiapina C. C. Pinheiro[2] · Plácido Rogério Pinheiro[1,2] · Luciano Comin Nunes[1,3] · Mirian Calíope Dantas Pinheiro[1] · Maria Lúcia Duarte Pereira[2] · Wilson Correia de Abreu[4] · Raimir Holanda Filho[2] · Marum Simão Filho[1,3] · Pedro Gabriel C. D. Pinheiro[1] · Rafael Espíndola Comin Nunes[3]

Received: 5 June 2021 / Accepted: 30 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract
Artificial Intelligence techniques based on Machine Learning algorithms, Neural Networks and Naïve Bayes can optimise the diagnostic process of the SARS-CoV-2 or Covid-19. The most significant help of these techniques is analysing data recorded by health professionals when treating patients with this disease. Health professionals' more specific focus is due to the reduction in the number of observable signs and symptoms, ranging from an acute respiratory condition to severe pneumonia, showing an efficient form of attribute engineering. It is important to note that the clinical diagnosis can vary from asymptomatic to extremely harsh conditions. About 80% of patients with Covid-19 may be asymptomatic or have few symptoms. Approximately 20% of the detected cases require hospital care because they have difficulty breathing, of which about 5% may require ventilatory support in the Intensive Care Unit. Also, the present study proposes a hybrid approach model, structured in the composition of Artificial Intelligence techniques, using Machine Learning algorithms, associated with multicriteria methods of decision support based on the Verbal Decision Analysis methodology, aiming at the discovery of knowledge, as well as exploring the predictive power of specific data in this study, to optimise the diagnostic models of Covid-19. Thus, the model will provide greater accuracy to the diagnosis sought through clinical observation.

Keywords Covid-19 · Machine learning · Verbal decision analysis · Hybrid model · Medical diagnostic optimization · Decision support systems

1 Introduction

The Covid -19 pandemic has increased dramatically since the first reported cases in China in December 2019 (Aggarwal 2018; World Health Organization 2020). Since then, more than 160 million cases of Covid-19 have been reported worldwide, with characteristics of Severe Acute Respiratory Syndrome by Coronavirus two infection (SARS-CoV-2). They were also causing more than 3.3 million deaths. The news in this context is that more

✉ Plácido Rogério Pinheiro
placido@unifor.br

Evandro Carvalho de Andrade
contact@evandroandrade.com

Luana Ibiapina C. C. Pinheiro
luanaibiapina1@hotmail.com

Luciano Comin Nunes
lcominn@uol.com.br

Mirian Calíope Dantas Pinheiro
caliope@unifor.br

Maria Lúcia Duarte Pereira
luciad029@gmail.com

Wilson Correia de Abreu
wjabreu@esenf.pt

Raimir Holanda Filho
raimir@unifor.br

Marum Simão Filho
marumsimao@gmail.com

Pedro Gabriel C. D. Pinheiro
pedrogcdpinheiro@hotmail.com

Rafael Espíndola Comin Nunes
rafael23.uni7@gmail.com

1 Graduate Program in Applied Informatics, University of Fortaleza, Fortaleza, Ceará, Brazil

2 Ceara State University, Fortaleza, Ceara, Brazil

3 University Center September 7, Fortaleza, Ceara, Brazil

4 School of Nursing of Porto, Porto, Portugal

Published online: 06 February 2023

than 1.38 million people in the world have already been vaccinated (Johns Hopkins University2021). The disease spectrum can range from asymptomatic infection to severe pneumonia with Acute Respiratory Distress Syndrome (ARDS) and death. The estimated incubation period is up to 14 days from the time of exposure, with an average incubation period of 4–5 days (Cao et al. 2020; Kucharski et al. 2020; Lauer et al. 2020). Anyone who has symptoms consistent with Covid-19 should be tested for SARS-CoV-2 infection (Liang 2020).

In a report on more than 373,000 confirmed Covid-19 cases with reported symptoms in the United States, 70% of patients experienced fever, cough, shortness of breath, 36% had muscle aches, and 34% reported headaches (Stokes et al. 2020). Other reported symptoms have included, but are not limited to, diarrhoea, dizziness, rhinorrhea, anosmia, dysgeusia, sore throat, abdominal pain, anorexia, and vomiting. More accurate clinical diagnoses and more effective treatments have driven computational methodologies based on Artificial Intelligence (A.I.). Activities such as remote monitoring, telemedicine, symptom association, epidemiological status alerts, exchange of clinical information, analysis of test results and treatment recommendations can use the A.I. The ability of machines to perform complex tasks and make decisions independently can improve outcomes be more efficient in establishing diagnoses and implementing treatments.

Algorithms based on Machine Learning (ML), one of the fields of A.I. study, have transformed medicine and its use, resulting in benefits for patients, doctors, and healthcare managers. Artificial Intelligence can help guide medical analysis, aiming to solve decision-making problems more realistically using relevant criteria and evaluation of several symptoms and risk factors, such as those existing in the cases of Covid-19. The early detection of the disease is essential to prevent the spread of the SARS-CoV-2 and the correct treatment in the initial phase, avoid the evolution of the cases, perform systematic monitoring in each stage of the disease, and improve the chances of hospital discharge. It is essential to highlight that the clinical diagnosis can vary from asymptomatic infections to extremely severe conditions. Within the database of the present study, approximately 80% of patients with Covid-19 were asymptomatic or had mild symptoms. Of the 20% of cases that presented moderate and severe symptoms, 43% required hospital care due to breathing difficulties, and 56% needed ventilatory support inside or outside the ICUs.

Additionally, to diagnose Covid-19, the association of the following signs and symptoms should be evaluated: fever, cough, dyspnea, runny nose, respiratory distress, sore throat, loss, or reduction of smell and or taste. Other risk factors are also observed, such as chronic cardiovascular diseases, myalgia, gastrointestinal disorders, diabetes mellitus and dyspnea. Its most common transmission is contact with another person who already has the virus through droplets of saliva and respiratory secretions.

Given the above, it is essential to emphasise that the present study proposes a hybrid approach model, structured in the juxtaposition of Artificial Intelligence techniques associated with the methodology of multicriteria decision analysis (MCDA). Regarding A.I., Machine Learning algorithms were used by the study. The choice for ML was due to its great potential in generating knowledge from large databases (Andrade et al. 2021b). In turn, about the MCDA methodology, the study was based on Verbal Decision Analysis methods, such as Ordinal Classification (ORCLASS) and ZAPROS-IIIi (Andrade et al. 2021a). The use of VDA is related to the need for decisions that involve subjective knowledge used in qualitative choices. Thus, the model proposed in this research aimed to discover expertise and explore the predictive power of study-specific data from a large set of well-defined information to optimise Covid-19 diagnostic protocols. In addition, the model provides greater accuracy to the diagnosis sought through clinical observation as a result.

The article is organised as follows: in addition to the Introduction Section, Sect. 2 presents details of the methods that make up the proposed model of this study. Next, the relationship between symptoms and diagnoses of the International Classification of Diseases (ICD) is presented in Sect. 3. In turn, in Sect. 4, the application of Verbal Decision Analysis methods is explained in detail. In Sect. 5, the analysis and results of the study are discussed. Finally, in Sect. 6, the conclusions are presented, and some future works related to the theme of this research are proposed. Additionally, the study shows the research limitations in the brief Sect. 7.

2 The model structure proposed

The construction of the proposed model used in this research used Verbal Decision Analysis (VDA) methods associated with Machine Learning (ML) techniques. Through VDA, the main criteria and respective alternatives were defined to formulate a questionnaire submitted to health professionals, aiming at a valuable validation of the model. Figure 1 shows the operation process of the model proposed for this study. In line with similar research, (Andrade et al. 2021a) conducted a study to obtain the main characteristics for the diagnosis of autism, as well as proposed a comparative study between ML algorithms and a hybrid model with Verbal Decision Analysis applied to the diagnosis of autism and optimisation of medical expertise (Andrade 2020).

The present research used the quantitative research method based on the responses to the questionnaire on Covid-19, which had been submitted to evaluate health

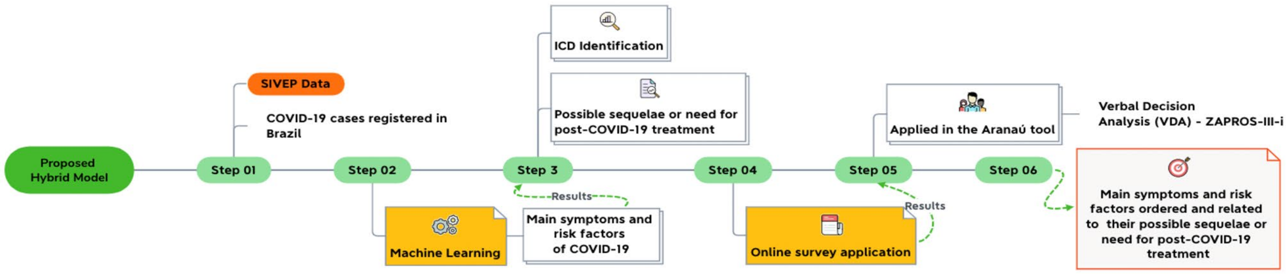

Fig. 1 Process of functioning of the proposed hybrid model structure

professionals. In addition, a qualitative approach was used to interpret the results of the mentioned assessments in the context of the effects of treatments applied to patients infected with SARS-CoV-2 in hospitals in Brazil. In the following subsections, there is a summary of the methods and techniques used in this research.

2.1 Machine learning

Considered a subset of Artificial Intelligence, Machine Learning allows a machine to have complete autonomy and control to learn on its own from previous data. Tom M. Mitchell, the principal exponent in ML, explains: "Machine learning is the study of computer algorithms that allow computer programs to improve with the experience" (Turban et al. 2011; Mitchell 2011) automatically. Machine Learning leads to Artificial Intelligence and depends on working with data sets, examining common patterns, and exploring complaints (Beğenilmiş and Uskudarl 2018; Mitchell 1997). The learning of the algorithms happens autonomously over time, according to the analysis of the training data. Its principal function is to guide computers to learn independently to improve their performance in the face of specific problems (Wilmott 2019).

Moreover, Machine Learning can recognise and extract patterns from a large volume of data with its learning characteristics, thus building a learning model. This learning is based on observing data, such as examples, direct experience or instruction. Once learning is done, it is possible to perform complex and dynamic tasks, predict more accurately, react in different situations, and behave intelligently (Delsentoth et al. 2020). According to well-defined objectives and known results, this study chose to use the Supervised Machine Learning methods to obtain knowledge and classification, identify trends and even predict outcomes based on historical data.

2.2 Machine learning algorithm considerations: logistic regression

Logistic Regression (L.R.), also known in the literature as logit regression, maximum-entropy classification (MaxEnt) or log-linear classifier, is a Machine Learning algorithm for classification problems and predictive analysis based on the concept of probability and used to assign observations to a discrete set of classes (Beğenilmiş and Uskudarl 2018; Hosmer and Lemeshow 2013). L.R. is one of the most popular AM algorithms for binary classification, given a set of independent variables to predict a binary result (1/0, Yes/No, True/False). It has simple complexity, and its application is successful in several areas, such as medicine, finance, and economics. In this model, the probabilities that describe the possible results of a single trial are modelled using a logistic function: a logical relationship between a dichotomous response variable and a series of numerical explanatory variables (continuous, discrete) "and"/"or" categorical.

Logistic Regression is a type of linear regression, when the result variable is categorical, in which the log of probabilities is used as a dependent variable. In simple words, it predicts the probability of an event occurring by adjusting data to a logit function (Hilbe 2016; Pant 2019). The algorithm makes various assumptions, such as independence responses (logits) at all subpopulation levels, where the explanatory variable is usually distributed. The variance is constant between the reactions and all the explanatory variable values. Intuitively, a transformation for the response variable is applied to produce a continuous probability distribution over output classes bounded between 0 and 1; this transformation is called the "logistic" or "sigmoid" function, where "z" corresponds to the logarithmic probabilities divided by the logit (Hoffman 2021). A logistic function or logistic curve is a standard S-shaped curve (sigmoid curve) with the equation (Bacaer 2011):

$$f(\boldsymbol{x}) = \frac{\boldsymbol{L}}{1 + \boldsymbol{e}^{-k(x-x_0)}}$$

$\boldsymbol{x}_0$ = x value of the sigmoid midpoint,$\boldsymbol{L}$ = maximum value of the curve,$\boldsymbol{k}$ = the logistic growth rate or slope of the curve.

L.R. is more of a classifier than a regression technique despite the name. Also, it is especially true in the scientific field, in medicine or the psychosocial sciences, where the focus is on prediction and explanation (Shmueli 2010).

2.3 Machine learning algorithm considerations: random forest

The random forest consists of random decision trees capable of predicting the expected values of regressions and classifications. Since its publication in 1996, the Breiman model (Breiman 1996) has become an important data analysis tool, quite versatile and with few adjustment parameters. The method is generally recognised for its precision and ability to handle small sample sizes, large resource spaces and complex data structures (Scornet et al. 2015). Several practical issues have successfully involved Random Forest, for example, air quality forecasting, chemoinformatics, ecology, 3D object recognition and bioinformatics. Many have proposed valid variations on the original algorithm to improve calculation times while maintaining good forecasting accuracy. Breiman's Random Forests were also extended to quantile estimation, survival analysis, and ranking forecast (Scornet et al. 2015). Theoretically, the analyses are less conclusive, and, regardless of their extensive use in practical contexts, little is known about the mathematical properties of Random Forests. Most studies have focused on isolated parts or simplified versions of the procedure. The most famous theoretical result is Breiman (1996), which offers an upper limit to the error of generalising forests in terms of the correlation and strength of individual trees. A critical step proposed by Jeon and Lin established lower limits for non-adaptive forests, independent of the training set (Jeon and Lin 2006).

The main computational procedures embedded in this algorithm are bagging and the criterion for classification and regression trees (CART) (Breiman et al. 1985). Bagging consists of a general aggregation scheme that generates subsamples of the original data set, building a predictor for each resampling and deciding on the mean. It improves unstable estimates, especially for extensive data and large data sets. In turn, the CART algorithm and its CART-split derivative are used to construct individual trees to choose the best cuts perpendicular to the axes. The best amount is selected in each tree node, optimising the CART split criterion based on the notion of Gini impurity (classification) and quadratic regression error (Breiman et al. 1985). Scornet, Biau and Vert studied some asymptotic properties of the Breiman algorithm in additive regression models. They proved the consistency of Random Forest, which provides a first basic theoretical guarantee of efficiency for this algorithm. This finding was the first consistent result for Breiman's original procedure. The approach was based on a detailed analysis of the cell behaviour generated by selecting the CART-split as the sample size increased. The study also showed that Random Forest could adapt to a sparse structure when the dimension is large, but only a smaller number of coordinates makes the information.

The general structure consists of the regression in which a random input vector X∈[0, 1]^p is observed—with "p" being the dimension of the vector. The objective is to predict the integrable random response square Y ∈ R by estimating the function of regression $m(x) = E[Y|X = x]$. Therefore, it is assumed that the training sample is given by $D_n = (X_1, Y_1), \dots, (X_n, Y_n) de[0,1]^p \times R$, independently distributed as the independent pair (X, Y). The objective is to use the data set D_n to construct an estimate $m_n : [0,1]^p \to R$ of the function "m". Thus, it is said that an estimated regression function m_n is consistent if $E\left[m_n(X) - m(X)\right]^2 \to \infty$ where the expectation E is more significant than X and D_n (Scornet et al. 2015). A random forest is a predictor that consists of a collection of M trees of random regression. For the j-th family tree, the predicted value at query point x is indicated by, $m_n(x, \Theta_j, D_n)$, where $\Theta_1, \dots, \Theta_M$ are independent random variables, distributed as a generic random variable Θ and independent of D_n. The trees are combined to form the finite estimate of the forest (Scornet et al. 2015).

$$m_{M,n}(x, \Theta_1, \dots, \Theta_M, D_n) = \frac{1}{M} \sum_{j=1}^{M} m_n(x, \Theta_j, D_n) \tag{1}$$

Since one can choose, in practice, M as large as possible, Scornet, Biau, and Vert (2015) demonstrated the property according to which the infinite forest estimate obtained as the limit of Eq. (1) is verified as follows when the number of M trees grows to infinity (Scornet et al. 2015):

$$m_n(x;D_n) = E_\Theta\left[m_n(x, \Theta, D_n)\right]$$

where E_Θ denotes expectation concerning the random parameter Θ, conditional on D_n. The law of large numbers, which states that it almost certainly depends on D_n, justifies this operation (Scornet et al. 2015):

$$\lim_{M\to\infty} m_{n,M}(x;\Theta_1, \dots, \Theta_M, D_n) = m_n(x;D_n);$$

See Breiman (Breiman 1996) for more details. From now on, to simplify the notation, $m_n(x)$ will be written instead of $m_n(x;D_n)$ (Scornet et al. 2015). In the original forests of Breiman (1996), each node of a single tree is associated with a hyper rectangular cell. At each stage of the tree's construction, the set of cells forms a partition [0, 1]^p. The tree's root is $[0, 1]^p$ itself, and each tree grows, as explained in Algorithm 1. This algorithm has three parameters (Scornet et al. 2015):

(1) $m_{try} \in \{1, ..., p\}$, which is the number of preselected directions for the tree's dividing.
(2) $a_n \in \{1, ..., n\}$, which is the number of data points sampled in each tree.
(3) $t_n \in \{1, ..., a_n\}$, which is the number of leaves in each tree.

By default, in the original procedure, the parameter $m_{tinytry}$ is set to p/3, a_n is set to n (resampling is done with substitution) and $t_n = a_n$. However, resampling is done without replacing this approach, and the parameters a_n and t_n may differ from their default values (Scornet et al. 2015). The algorithm works by growing M different trees as follows: For each tree, data points are drawn at random without replacing the original data set; then, in each cell of each tree, a division is chosen maximising the CART criterion; finally, the construction of each tree is interrupted when the total number of cells in the tree reaches the value t_n. Therefore, each cell contains exactly one point in the case $t_n = a_n$ (Scornet et al. 2015).

2.4 Machine learning algorithm considerations: naïve bayes

Bayesian networks can be applied to classification problems using Bayesian classifiers. The simplest Bayesian Network structure constitutes a classification algorithm called Naïve Bayes, also known as the Bayesian classifier with k-dependency and the structure of Simple Bayesian Networks with K-Dependency (Castro et al. 2011; Sahami 1996). Therefore, Naïve Bayes is a simple probabilistic algorithm based on Bayes' theorem. This algorithm uses training data to form a probabilistic model based on the evidence of specific characteristics in an extensive database. The algorithm assumes independence between the model's features, implying the absence of relations between a given quality and the other attributes in the database. A Naïve Bayes classifier is structured in the calculation of the posterior probability distribution P(Y|X), where $\boldsymbol{Y} = \left(y_1, y_2, \dots y_k\right)$ is the random variable to be classified presenting "k" categories and $\boldsymbol{X} = \left(x_1,\ x_2,\ \dots, x_p\right)$ is a set of "p" discrete explanatory variables. For the calculation of conditional probability P(Y|X), this method assumes probabilistic independence between the explanatory variables, facilitating the application of the technique computationally, the result of this calculation being obtained by the following formula:

$$P\left(Y = y_k | x_1, \dots, x_p \propto P\left(Y = y_k\right) \prod_{i=1}^{p} P\left(x_i | Y = y_k\right) \right)$$

Therefore, a Naïve Bayes classifier calculates the probability that a given observation belongs to each category and classifies that observation in the most appropriate variety.

2.5 Machine learning algorithm considerations: neural networks

Artificial neural networks are Machine Learning techniques that simulate learning in biological organisms. The human nervous system contains cells, which are referred to as neurons. Synapses connect the neurons. The strengths of synaptic connections often change in response to external stimuli. This change is how learning takes place in living organisms. This biological mechanism is simulated in artificial neural networks containing neurons or "perceptrons" units. Neurons are connected employing weights, which play the same role as synaptic connections in biological organisms.

Neural Networks have some advantages over traditional Machine Learning. The first advantage is that Neural Networks provides a top-level abstraction to express semantic insights about data domains through architectural design options in the computational graph. The second advantage is that Neural Networks offer a simple way to adjust the complexity of a model, adding or removing neurons from the architecture according to training data or computational power availability. In the recent success of Neural Networks, the increased availability of data and the processing power of modern computers have exceeded the limits of traditional Machine Learning algorithms. The performance of conventional Machine Learning sometimes remains better for smaller data sets due to more options, more accessible interpretation of the model under analysis and the tendency to create hand-interpretable resources incorporating domain-specific perceptions. With limited data, the best of a wide variety of models in Machine Learning will generally outperform a single class of Neural Networks models.

As for the type of learning, Neural Networks can be categorised as follows: Supervised Learning, Unsupervised or Hybrid Learning.

2.6 Verbal decision analysis methods

Verbal Decision Analysis (VDA) is an approach introduced in 1997 by Larichev and Moshkovich (1997, 1994, 1995). VDA seeks to express the decision-maker's assessments and preferences verbally and qualitatively. In this approach, natural language forms the questions, and verbal evaluations include the scales of criteria (Machado et al. 2014).

Table 1 Primary symptoms and risk factors identified by logistic regression source: (Andrade et al. 2021b)

Principal symptoms of COVID-19		Principal risk factors of COVID-19
Saturation below 95%	Vomit	Immunodeficiency
Sore throats	Respiratory discomfort	Chronic kidney disease
Loss of smell	Cough	Diabetes mellitus
Fatigue	Fever	Chronic cardiovascular disease
Abdominal pain		Dyspnoea

2.6.1 The ZAPROS method family

ZAPROS consists of Russian words meaning "locked procedures close to reference situations". The first version added the creators' first names initials "L.M." (Larichev and Moshkovich) to the name, forming ZAPROS-LM (Larichev and Moshkovich 1995). This method is destined to aid in resolving problems that need to classify and sort a large number of alternatives with sets capable of changing the decision rules. The ZAPROS method presents the decision-maker's preferences by comparing hypothetical alternatives pairs of any set close to so-called "landmarks": the better or worse possible values. These alternatives compose only two criteria, and the decision-maker must indicate the preference for one of the indifferences among both.

O method evaluates information's consistency through the preferences' transitivity and compares it to the same criteria values at two landmarks. As a result of these comparisons' complete set, the Joint Ordinal Scale (JOS) is formed in the scope of the criteria. This scale sorts the different criteria and is not linked to a specific set of alternatives. Therefore, the Joint Ordinal Scale (JOS) enables partial comparisons of natural alternatives pairs. These form the basis for partisan sorting of valid alternatives and constitute the ZAPROS method's focus (Larichev and Moshkovich 1995).

2.7 Orange canvas framework considerations

The present study used the Orange Canvas framework to obtain the best possible results when comparing the algorithms referenced in this work. Demsar et al. (2013) designed a framework called Orange, whose structure consists of a set of machine learning and data mining tools for data analysis using Python coding language scripts, in addition to visual programming. Orange is aimed at experienced users and programmers, students, and data scientists who constantly use data mining techniques. This framework is based on the C++ language; however, it allows code developers to work with Python. In addition, (Andrade et al. 2021b) used Machine Learning (ML) algorithms and their concepts through the Orange tool in a data volume of 51,560 (fifty thousand five hundred and sixty) Covid-19 cases registered in Brazil by "SIVEP Influenza—Sistema de Epidemiological Surveillance of Influenza". This study classified the main symptoms and risk factors from a minor symptomatic perspective, identified by the Logistic Regression algorithm when applied in the SIPEV database. Table 1 presents these main symptoms and risks.

Table 2 Results—cases of hospital discharge

Model	Train Time(s)	Test time (s)	Classification Accuracy	Precision
LR	5.395	0.357	**0.615**	**0.604**
NN	186.241	0.732	0.595	0.576
RF	2.118	0.604	0.587	0.564
NB	0.319	0.127	0.181	0.317

Table 3 Results—cases of death

Model	Train time (s)	Test time (s)	Classification Accuracy	Precision
LR	8.073	0.401	0.622	0.585
RF	2.594	0.661	0.576	0.550
NN	230.880	0.900	0.571	0.553
NB	0.502	0.079	0.311	0.529

Using the Orange tool, with the same data as the study to obtain symptoms and risk factors, the Neural Networks algorithm was added to analyse its performance concerning the others used in the previous study.

2.8 Performance analysis

The authors made some comparisons using the software Orange, between the Logistic Regression (L.R.), Random Forest (R.F.), Naïve Bayes (N.B.), and Neural Network (N.N.) algorithms to identify which could improve, for example, in the list of symptoms or even in the performance and precision in obtaining this data. The proposed model was refined from several test bases, one for each learning cycle. The origin of each test base of each cycle counted on the exploration of a large and varied amount of data structured in rows and columns. In this way, it is understood that there is no trend bias in the model, given the magnitude and variety of the data source for each test base. In addition to the test cycles, there is a horizon based on the experience and common sense of the professionals who validated the model. The metrics presented in Tables 2 and 3 show the best performing Logistic Regression algorithm in almost all measures, concerning Neural Network, Random Forest, and Naïve Bayes. Additionally, it offers a better Classification Accuracy (C.A.,) indicating that many are correctly classified among all the classifications of Covid-19 cases performed.

Finding an algorithm with better performance is very relative, according to the idea of The No Free Lunch theorem for supervised algorithms (Wolper April 1997). After completing the comparison of the algorithms, we chose to follow the studies with the same symptoms and risk factors previously classified by Logistic Regression, Table 2, because even in comparison with Neural Networks (algorithm included in this study), it presented the best accuracy and precision above the studied data.

3 The relationship: symptoms, international classification of disease and possible sequelae

The International Classification of Disease (ICD) was obtained through the identified symptoms and risk factors, referring to each (ICD 2019). Based on these ICDs, the authors prepared a table, with the help of a health professional, containing the symptoms, risk factors, their related ICDs and the possible sequelae or need for post-Covid-19 treatment, as shown in Fig. 2 and Fig. 3.

3.1 Application phase

Through the identified symptoms and risk factors, the International Classification of Disease (ICD) was obtained from the collection of data from Fig. 3 and Fig. 4. The evaluation criteria were defined according to the chapter of the manual of the International Classification of Functioning, Disability and Health—ICF (World Health Organization 2001), using the broadest classification of the category (chapter), as shown in Table 4. Each criterion is registered to consider the recurrence of the symptom or risk factor, with possible sequelae or need for treatment related to the analysed criterion. Also, Table 4 below shows the qualifiers that specify this recurrence.

Health professionals created an online survey to validate the symptoms identified by the machine learning algorithm

	Symptoms	ICD	ICD - Description	ICD - Chapter	Possible Sequels
1	Saturation Below 95%	J96 J96.0 J96.1	Respiratory failure, not elsewhere classified Acute respiratory failure Chronic respiratory failure	Chapter X Diseases of the respiratory system (J00-J99)	1 - Long-term Damage To Your Lungs / 2 - Pulmonary Rehabilitation, Which Includes Exercise Therapy, Education, And Counseling. / 3 - Continuous Help With Your Breathing / 4 - Develop An Abnormal Heart Rhythm / 5 - Stop Breathing / 6 - Slip Into A Coma. 7 - Pulmonary Fibrosis
2	Sore Throats	R07	Pain in throat and chest	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Pneumonia
3	Loss of Smell	R43 R43.0	Disturbances of smell and taste Anosmia	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Sensitivity Changes
4	Fatigue	R53	Malaise and fatigue	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Strength / 2 - Muscle Weakness / 3 – Depression / 4 – Anxiety / 5 - Cognitive Changes
5	Abdominal Pain	R10 R10.1	Abdominal and pelvic pain Pain localized to upper abdomen	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Being Prostrate / 2 - With Diarrhea / 3 - Not Being Able To Drink Water / 4 - Not Being Able To Eating Right
6	Vomit	R11	Nausea and vomiting	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Sudden And Severe Headache With No Apparent Cause / 2 - Loss Of Vision / 3 - Sudden Vertigo / 4 - Imbalance
7	Respiratory Discomfort	U04 U04.9	Severe acute respiratory syndrome [SARS] Severe acute respiratory syndrome [SARS], unspecified	Chapter XXII Codes for special purposes (U00-U85)	1 - Reduced Lung Capacity
8	Cough	R05	Cough	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Residual Cough
9	Fever	R50	Fever of other and unknown origin	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Convulsions

Fig. 2 Symptoms and possible sequels

	Risk Factors	ICD	ICD - Description	ICD - Chapter	Possible Sequels
1	Immunodeficiency	D80	Immunodeficiency with predominantly antibody defects	Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89)	1 - High Risk Of Toxoplasmosis / 2 – Encephalitis / 3 - Cryptococcal Meningitis / 4 - Cytomegalovirus / 5- Lymphoma
		D81	Combined immunodeficiencies		
		D82	Immunodeficiency associated with other major defects		
		D83	Common variable immunodeficiency		
		D84	Other immunodeficiencies		
2	Chronic Kidney Disease	N18	Chronic kidney disease	Chapter XIV Diseases of the genitourinary system (N00-N99)	1 - Decreases The Body's Ability To Fight Infections. / 2 – Anemia / 3 - Gastric Or Intestinal Bleeding / 4 - Pain In Bones / 5 - Pain In Joints / 6 - Pain In Muscles / 7 - Metabolic Acidosis / 8 - Malnutrition And Alteration Of Calcium And Phosphorus Metabolism
		N18.9	Chronic kidney disease, unspecified		
		N19	Unspecified kidney failure		
3	Diabetes Mellitus	E10	Type 1 diabetes mellitus	Chapter IV Endocrine, nutritional and metabolic diseases (E00-E90)	1 - Imbalance Of The Entire Balance Of Body Functions, Including An Increase In Blood Glucose / 2 - Lack Of Blood Glucose Control / 3 - Stressful Conditions Lead To High Levels Of Regulatory Hormones That Increase Blood Sugar To Help The Body Fight Infection
		E11	Type 2 diabetes mellitus		
		E12	Malnutrition-related diabetes mellitus		
		E13	Other specified diabetes mellitus		
		E14	Unspecified diabetes mellitus		
4	Chronic Cardiovascular Disease	I51.6	Cardiovascular disease, unspecified	Chapter IX Diseases of the circulatory system (I00-I99)	1 - Cardiac Sequelae / 2 - Imbalance Between High Metabolic Demand And Low Reserve Cardiac Arrest / 3 - Systemic Inflammation / 4 - Thrombogens And Cardiac Injury direct By The Virus
		I51.9	Heart disease, unspecified		
5	Dyspnoea	R06.0	Dyspnoea	Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99)	1 - Shortness Of Breathe / Difficulty Breathing

Fig. 3 Risk factors and possible sequels

and the sequelae (Andrade et al. 2021b). Fourteen possible relationships between symptoms and sequelae were addressed in the questionnaire. The professional qualified the questions based on the criteria and qualifiers (recurrence) when answering the questions and assessing them. Twenty individual specialists working in the health area were interviewed. Nurses, physiotherapists, doctors, occupational therapists, phono audiologists, psychologists, dentists, pharmacists, and researchers responded to the survey in the interview.

In addition to collecting data that meets the concept, obtaining it is extremely important so that the necessary treatment can be applied later to test the hypotheses. Therefore, it is essential to pay attention to the design of the collection instrument, the type of information it will provide and the type of analysis that can be done after obtaining the responses to the form. From the data collection of useful or pertinent data for testing the hypotheses, the analysis model was compared with the collected data, considering the requirements to correctly put the chosen method into practice and obtaining elements to achieve the objectives proposed in this research. After obtaining the survey responses, the result was applied to the Aranau tool that implements ZAPROS-IIIi.

4 Application of the verbal decision analysis methods

The results were obtained through the hybrid model using Machine Learning, with Logistic Regression and the ZAPROS-IIIi method of the Verbal Decision Analysis methodology, to determine the relationship between symptoms/risk factors and possible sequelae/treatment. After the application phase, in which the evaluations of each professional were obtained through responses to the online questionnaire, a table was created with the 14 (fourteen) symptoms/risk factors and their respective relationship alternatives (low, moderate, and high). The most significant response is highlighted for each relationship between symptom/risk factor with possible sequelae or need for treatment after the disease, as shown in Fig. 4. For example, each question was evaluated, the first question: Regarding criterion A, fourteen professionals answered that there might be a high relationship between the symptom and the sequel under the analysed criterion. Likewise, all research questions and their answers were analysed.

According to the evaluation carried out for each question, the result was recorded in the Aranau tool. Figure 5 exemplifies the first 3 (three) alternatives to be evaluated by the decision-maker, using the Aranau tool, which implements ZAPROS-IIIi. On the other hand, following the first question's application, it was included in the tool according to the number of responses for each criterion and qualifier: A1—High / B1—High/C2—Moderate.

After defining the relationships, Elicitation of Preferences was initiated. The qualifiers A1 (High), B1 (High) and C1

Research Project: COVID-19 and possible sequelae / need for treatments in the post-disease period.				
QUESTIONS	CRITERIA	ALTERNATIVES		
		Low	Moderate	High
1. The patient presented the saturation symptom below 95% during treatment against COVID-19 and is currently observed in his clinical picture: possible long-term lung injury, abnormal heart rhythm, pulmonary fibrosis, need for pulmonary rehabilitation (physical therapy , education and counseling) and ongoing help to develop breathing.	A	2	4	14
	B	4	6	10
	C	8	9	3
2. The patient had a sore throat symptom during treatment against COVID-19 and is currently seen in his clinical picture: pneumonia.	A	3	5	12
	B	4	7	9
	C	11	8	1
3. The patient presented the symptom of loss of smell during treatment with COVID-19 and is currently observed in his clinical condition: alteration of sensitivity.	A	7	10	3
	B	7	10	3
	C	5	6	9
4. The patient presented the symptom of fatigue during treatment against COVID-19 and is currently observed in his clinical condition: lack of strength, muscle weakness, depression, anxiety and cognitive changes.	A	2	4	14
	B	2	7	11
	C	2	8	10
5. The patient presented the symptom of abdominal pain during treatment with COVID-19 and is currently observed in his clinical condition: in a state of prostration, with diarrhea, not being able to drink water and eat properly.	A	11	6	3
	B	6	8	6
	C	10	10	0
6. The patient presented the symptom of vomiting during treatment with COVID-19 and is currently observed in his clinical condition: sudden and severe headache without apparent cause, loss of vision and sudden vertigo.	A	8	9	3
	B	6	8	6
	C	6	8	6
7. The patient presented the symptom of difficulty breathing during treatment with COVID-19 and is currently observed in his clinical condition: reduced lung capacity.	A	1	4	15
	B	3	7	10
	C	8	9	3
8. The patient presented the cough symptom during treatment with COVID-19 and is currently observed in his clinical condition: residual cough.	A	2	7	11
	B	6	7	7
	C	10	8	2
9. The patient presented the feverish symptom during treatment with COVID-19 and is currently in his clinical condition: seizures.	A	6	3	11
	B	5	7	8
	C	5	5	10
10. The patient with an immunodeficiency risk factor, was treated against COVID-19 and is currently seen in his clinical condition: high risk of toxoplasmosis, encephalitis, cryptococcal meningitis, cytomegalovirus and lymphoma.	A	6	9	5
	B	4	4	12
	C	4	8	8
11. The patient with the risk factor Chronic Kidney Disease during treatment against COVID-19 and currently has a clinical condition: decreased ability of the body to fight infections, anemia, gastric or intestinal bleeding, bone pain, joint pain, muscle pain, metabolic acidosis, malnutrition and alteration of calcium and phosphorus metabolism.	A	5	9	6
	B	2	8	10
	C	8	9	3
12. The patient with the risk factor Diabetes Mellitus during treatment against COVID-19 and currently observes in his clinical picture: imbalance of bodily functions, including increased blood glucose, lack of glycemic control (stressful conditions lead to high levels of regulating hormones that increase blood sugar to help the body fight infections).	A	2	12	6
	B	0	7	13
	C	5	10	5
13. Patient with the risk factor Chronic Cardiovascular Disease during treatment against COVID-19 and is currently observed in his clinical condition: cardiac sequelae, imbalance between high metabolic demand and low reserve, cardiac arrest, systemic inflammation and thrombogenesis.	A	1	7	12
	B	0	11	9
	C	7	11	2
14. Patient with the risk factor Dyspnea during treatment against COVID-19 and currently has a clinical condition: shortness of breath / difficulty breathing.	A	1	5	14
	B	4	7	9
	C	9	8	3

Fig. 4 Possible sequelae/need for treatments in the post-disease period

Table 4 Evaluation criteria

Evaluation criteria	Evaluation qualifiers
A—In functions of the respiratory/cardiovascular system	1. High: In most cases, the relationship between the symptom/risk factor presented is observed, with the sequelae/need for treatment listed
B—In functions of the haematological/immunological systems	2. Moderate: The relationship between the symptom/risk factor presented is observed in some cases, with the sequelae/need for treatment listed
C—In functions of the neurological system	3. Low: In most cases, there is no observed or no relationship between the symptom/risk factor presented with the sequelae/need for treatment listed

(High) are considered the ideal result, or the one that would have the most significant impact on the relationship between symptom X/risk factor X with the possible sequel Y or the need for post-treatment disease. Having defined the ideal reference situation, the ZAPROS-IIIi method, based on the concept of Quality Variations (Q.V), presents the minimum number of qualifying pairs needed to be compared by the decision-maker. The process uses the measurement of the distances between the evaluations of the two alternatives and the Formal Index Quality (F.I.Q). The Aranau tool presents the Q.V.s for comparison of the decision-maker. The question "What is the recurrence of sequel X or the need for treatment X to occur for a symptom Y?" is considered. The tool performed the sum of the assessments for each

Defined Alternatives

Alternatives	Values
1. The patient presented the saturation symptom below 95% during treatment against COVID19 and is currently observed in his clinical picture: possible long-term lung injury, abnormal heart rhythm, pulmonary fibrosis, need for pulmonary rehabilitation (physical therapy , education and counseling) and ongoing help to develop breathing.	A1 – High: In most cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	B1 – High: In most cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	C2 – Moderate: In some cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
2. The patient had a sore throat symptom during treatment against COVID19 and is currently seen in his clinical picture: pneumonia.	A1 – High: In most cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	B1 – High: In most cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	C3 – Low: In most cases, there is no observed or no relationship between the symptom / risk factor presented, with the sequelae / need for treatment listed.
3. The patient presented the symptom of loss of smell during treatment with COVID19 and is currently observed in his clinical condition: alteration of sensitivity.	A2 – Moderate: In some cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	B2 – Moderate: In some cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.
	C1 – High: In most cases, the relationship between the symptom / risk factor presented is observed, with the sequelae / need for treatment listed.

Fig. 5 Defined alternatives

The preferred order among the main symptoms / risk factors of COVID-19 with possible sequelae and / or need for treatments.

RANk	QUESTIONS	SYMPTOM / RISK FACTOR	SEQUEL	FIQ
1	4. The patient presented the symptom of fatigue during treatment against COVID19 and is currently observed in his clinical condition: lack of strength, muscle weakness, depression, anxiety and cognitive changes.	Fatigue	lack of strength, muscle weakness, depression, anxiety and cognitive changes.	0
2	9. The patient presented the feverish symptom during treatment with COVID19 and is currently in his clinical condition: seizures.	Feverish	Seizures	0
3	1. The patient presented the saturation symptom below 95% during treatment against COVID19 and is currently observed in his clinical picture: possible long term lung injury, abnormal heart rhythm, pulmonary fibrosis, need for pulmonary rehabilitation (physical therapy, education and counseling) and ongoing help to develop breathing.	Saturation symptom below 95%	Possible long term lung injury, abnormal heart rhythm, pulmonary fibrosis, need for pulmonary rehabilitation (physical therapy, education and counseling) and ongoing help to develop breathing	1
4	7. The patient presented the symptom of difficulty breathing during treatment with COVID19 and is currently observed in his clinical condition: reduced lung capacity.	Difficulty breathing	Reduced lung capacity	1
5	11. The patient with the risk factor Chronic Kidney Disease during treatment against COVID19 and currently has a clinical condition: decreased ability of the body to fight infections, anemia, gastric or intestinal bleeding, bone pain, joint pain, muscle pain, metabolic acidosis, malnutrition and alteration of calcium and phosphorus metabolism.	Chronic Kidney Disease	Decreased ability of the body to fight infections, anemia, gastric or intestinal bleeding, bone pain, joint pain, muscle pain, metabolic acidosis, malnutrition and alteration of calcium and phosphorus metabolism.	4
6	12. The patient with the risk factor Diabetes Mellitus during treatment against COVID19 and currently observes in his clinical picture: imbalance of bodily functions, including increased blood glucose, lack of glycemic control (stressful conditions lead to high levels of regulating hormones that increase blood sugar to help the body fight infections).	Diabetes Mellitus	Imbalance of bodily functions, including increased blood glucose, lack of glycemic control (stressful conditions lead to high levels of regulating hormones that increase blood sugar to help the body fight infections).	4
7	14. Patient with the risk factor Dyspnea during treatment against COVID19 and currently has a clinical condition: shortness of breath / difficulty breathing.	Dyspnea	Shortness of breath / difficulty breathing	4
8	13. Patient with the risk factor Chronic Cardiovascular Disease during treatment against COVID19 and is currently observed in his clinical condition: cardiac sequelae, imbalance between high metabolic demand and low reserve, cardiac arrest, systemic inflammation and thrombogenesis.	Chronic Cardiovascular Disease	Cardiac sequelae, imbalance between high metabolic demand and low reserve, cardiac arrest, systemic inflammation and thrombogenesis	3
9	10. The patient with an immunodeficiency risk factor, was treated against COVID19 and is currently seen in his clinical condition: high risk of toxoplasmosis, encephalitis, cryptococcal meningitis, cytomegalovirus and lymphoma	Immunodeficien cy	High risk of toxoplasmosis, encephalitis, cryptococcal meningitis, cytomegalovirus and lymphoma	4
10	2. The patient had a sore throat symptom during treatment against COVID19 and is currently seen in his clinical picture: pneumonia.	Sore throat	Pneumonia	4
11	3. The patient presented the symptom of loss of smell during treatment with COVID19 and is currently observed in his clinical condition: alteration of sensitivity.	Loss of smell	Alteration of sensitivity	5
12	6. The patient presented the symptom of vomiting during treatment with COVID19 and is currently observed in his clinical condition: sudden and severe headache without apparent cause, loss of vision and sudden vertigo.	Vomiting	Sudden and severe headache without apparent cause, loss of vision and sudden vertigo	6
13	8. The patient presented the cough symptom during treatment with COVID19 and is currently observed in his clinical condition: residual cough.	Cough	Residual cough	6
14	5. The patient presented the symptom of abdominal pain during treatment with COVID19 and is currently observed in his clinical condition: in a state of prostration, with diarrhea, not being able to drink water and eat properly.	Abdominal pain	In a state of prostration, with diarrhea, not being able to drink water and eat properly	12

Fig. 6 Preferred order

character, according to the comparison of the V.Q.s. The best alternatives obtained in the comparison received a lower Formal Index Quality. These data allowed the combination of vectors formed by the "Qualifiers X Alternatives", that is, "rows X columns" of Fig. 4. Figure 6, in turn, shows the final order of preference for the central relationships between symptoms/risk factors and their possible sequelae or need for post-disease treatment, established by the ZAPROS-IIIi method, implemented in the Aranau tool for the cases of COVID-19. Figure 7 shows this order graphically.

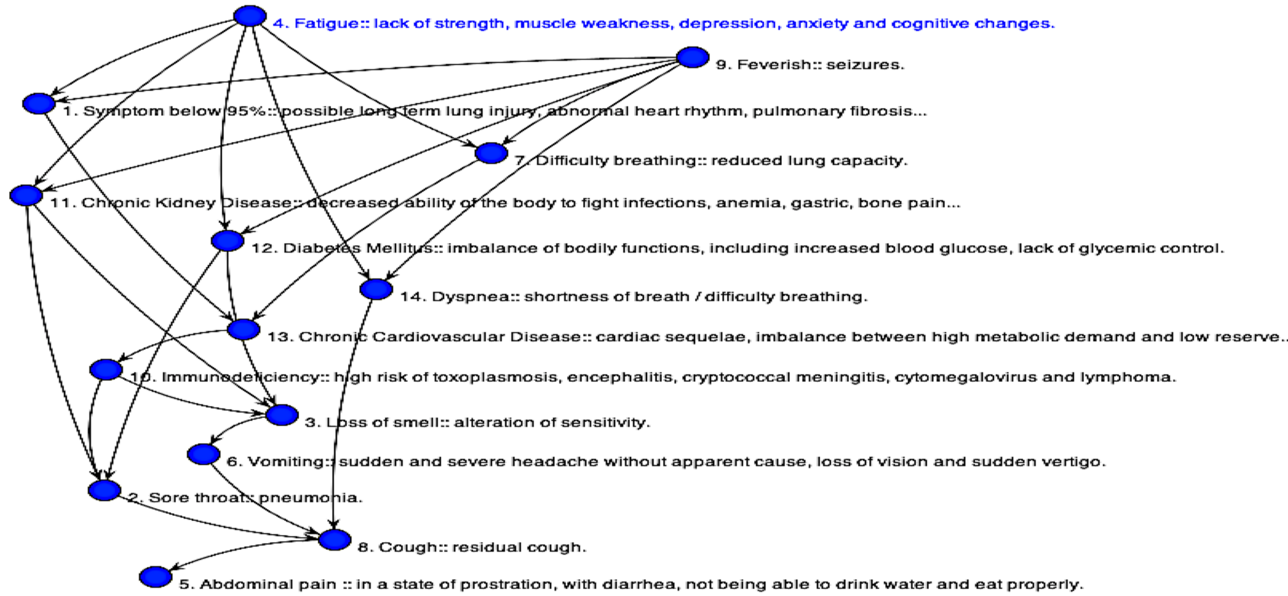

Fig. 7 Alternatives graph

5 Analysis and discussion of results

The post-COVID-19 syndrome is a set of symptoms that persisted in patients infected with SARSCOV-2, both in the manifestation of the disease in its mild/moderate form and its critical condition leading to hospitalisation for intensive long-term care infection (Irwin et al. 2018). Thus, patients surviving COVID-19 even after hospital discharge need immediate and long-term care related to psychomotor and social care. Persistent fatigue is one of the symptoms identified in the long term. When associated with depressive symptoms, anxiety, and diffuse myalgia, they are characterised as chronic fatigue syndrome (CFS), which can be caused due to the febrile reaction and the immunodepression associated with viral status (Perrin et al. 2020). According to research carried out by Carfi et al. (2020), among the symptoms analysed by patients who were discharged from hospital after the cure of COVID-19, 53.1% of the people still revealed to have cognitive dysfunction, anxiety, depression, in addition to fatigue and muscle weakness, due to prolonged hospitalisation in the Intensive Care Unit (ICU).

As predictors of persistent pulmonary fibrosis are the constant high inflammatory load, the PaO2/FiO2 ratio < 324 and Body Mass Index (BMI) > 33 kg/m^2 at hospital admission. The increase in anti-inflammatory cytokine (Interleukin-4) and macrophages are triggered by blocking the conversion of angiotensin-2 to angiotensin, inducing a pro-inflammatory state. The viral protein composed of a single RNA chain can be expressed in the bronchial and parenchymal pathways, corroborating lung injury and vascular endothelium, glial, renal, and enteric cells (Santana et al. 2021). The binding of the virus to the receptors of the Angiotensin-2 converting enzyme (ACE2) favours viral penetration into cardiac cells, triggering sequelae and a high risk of thrombogenesis, which persist after infection SARS-CoV2 (Santana et al. 2021). Studies indicate a high prevalence of COVID-19 patients with damage to the cardiovascular system, such as acute cardiac injury of 7.2%, cardiogenic shock 8.7% and cardiac arrhythmia of 16.7%, which were caused due to the systemic inflammatory response by the overproduction of inflammatory cytokines, IL-6 and TNF-α. Disorders of the immune system during the progression of the disease can lead to hypercoagulability, a high mortality rate and the need for cardiovascular protection during drug therapy for COVID-19 (Ferrari 2020).

After clinical discharge, due to the context of respiratory failure and immunological deficiency, hypoperfusion of the frontal lobe can be evidenced, and encephalopathy, which involves the rupture of the blood–brain barrier by the indirect invasion of the virus through cytokines (Mao et al. 2020). In addition to pulmonary involvement in patients with SARS-CoV-2, other organs can suffer complications such as the development of Acute Kidney Injury (AKI). Research infers those patients with dialysis Chronic Kidney Disease (CKD) affected by COVID-19 may develop severe immune dysregulation and exaggerated production of cytokines of acute infection, causing lymphopenia, anaemia, gastric or intestinal bleeding, bone and joint pain, myalgia, acidosis metabolic, malnutrition and alteration of the metabolism of calcium and phosphorus (Pecly et al. 2021).

Thus, the early detection of renal abnormalities is essential to ensure a better evolution of these patients during hospitalisation by planning excellent hemodynamic support when indicated and avoiding nephrotoxic drugs through medical risk judgment (Pecly et al. 2021). Another point to be discussed during the COVID-19 pandemic is the risk factor Diabetes Mellitus, a chronic disease capable of causing cellular and vascular metabolism changes. As a result, diabetic people are at high risk since hyperglycemia and insulin resistance that may be present in this public promote a more remarkable synthesis of glycation products, pro-inflammatory cytokines and oxidative stress capable of stimulating the adhesion of inflammatory mediators and carry a higher risk of pulmonary infection (Marinho et al. 2021). According to World Health Organization (WHO), the initial signs and symptoms of the disease are similar to the flu condition. They can vary from each individual, manifested in mild form as pneumonia and Severe Acute Respiratory Syndrome (SARS). Most infected people experience headaches early in the disease, which can persist due to activation of the nociceptive system by the inflammatory cascade; cough; sore throat; diarrhoea; abdominal pain; nausea and vomiting. Hyposmia is also a common symptom in COVID-19 in the first eight days of the disease since the viral contamination of the olfactory bulb is considered the gateway of the virus to the Central Nervous System, reaching the olfactory cortex and structures of the diencephalon. Thus, the SARS-CoV-2 invasion process can cause an axonal lesion, affecting the individual, even when cured, a change in sensitivity secondary to inflammatory microangiopathy of the nasal epithelium (Waheed et al. 2020).

6 Conclusions and future works

Complex problems are usually solved by breaking them into less complicated parts. Therefore, hybrid models are heavily used, as this facilitates the implementation of methodologies with their methods in each piece. The model proposed in the present study shows that this form of execution was perceived in stages of the decomposition of the problem studied. Thus, the hybrid approach model structured in the composition of Artificial Intelligence techniques (Machine Learning) associated with the Multicriteria Decision Analysis Methodology (Verbal Decision Analysis) ordered the main symptoms and factors of COVID-19 with their respective persistent changes after clinical discharge and cured the disease. With the application of Machine Learning, Logistic Regression was obtained as the best performing algorithm with an accuracy of 60.4%; Neural Network (57.6%), Random Forest (56.4%), Bayesian Naïve (31.7%). In turn, in the qualitative solution sought for the order of preference of the characteristics according to the Verbal Decision Analysis with the evaluation of specialist professionals in the first place, presented the symptom "fatigue" with the respective sequelae: lack of strength, muscle weakness, depression, anxiety and cognitive changes. Lastly, he showed the sign "abdominal pain", whose sequelae were: prostration, diarrhoea, and inability to drink water and eat properly. The research theme has shown outstanding interest according to Gupta et al. (2021); Verma and Rathi 2022).

Therefore, hybrid models, such as the one proposed in this research, provide the resolution of complex problems as described in the Introduction Section. Next, some future work and possible refinements presented in this research are suggested, in addition to some new ideas that emerged during the development of the study: seek to identify treatments for potential sequelae of the disease involved in the study; use machine learning algorithms to rank drugs with the best results for COVID-19; seek to use machine learning algorithms to organise the best treatments for the symptoms of said disease, listed in this study. Developing other hybrid models will help discover proactive solutions to prevent similar or even more severe pandemic events and combat them. Such models can use Data Science techniques based on Artificial Neural Networks, Deep Learning, Data Mining. All these techniques may be associated with methods from the MCDA family.

7 Limitations

The present study was limited to developing a protocol structured in a hybrid model of Machine Learning and Verbal Decision Analysis, identifying the possible sequelae or necessary treatments related to the main symptoms and risk factors of COVID -19.

Acknowledgements To all health professionals and diagnosed patients who fought or fought an arduous battle against COVID-19. The third, sixth and eighth authors are grateful to the Brazilian National Council for Scientific and Technological Development (CNPq).

Declarations

Conflict of interest We have no conflict of interest to declare.

References

Aggarwal CC (2018) Neural networks and deep learning. Springer International Publishing AG Springer Nature, New York

Andrade E (2020) Um Modelo Híbrido em Aprendizagem de Máquina e Análise Verbal de Decisão Aplicado ao Diagnóstico de Autismo, Masters Dissertation, University of Fortaleza

Andrade E, Portela S, Pinheiro, PR, Nunes LC, Simão Filho M, Costa, WS, Pinheiro MCD (2021a) A protocol for the diagnosis of autism spectrum disorder structured in machine learning and verbal decision analysis. Comput Math Methods Med 2021:1628959. https://doi.org/10.1155/2021/1628959

Andrade E, Pinheiro PR, Nunes LC, Pinheiro MCD, HolandaFilho R, Abreu W, SimãoFilho M, Pinheiro LICC, Pereira MLD (2021b) Application of machine learning to infer symptoms and risk factors of Covid-19, research, and innovation forum. Springer, New York

Bacaer N (2011) Verhulst and the logistic equation (1838), A short history of mathematical population dynamics. Springer, London, pp 35–39

Beğenilmiş E, Uskudarl S (2018) Organised behavior classification of tweet sets using supervised learning methods. In: Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS '18). ACM

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140. https://doi.org/10.1023/A:1018054314350

Breiman L, Friedman J, Olshen R, Stone EC (1985) Classification, and regression trees. The Wadsworth Statistics/Probability Series, Wadsworth

Cao Z, Li T, Liang LL, Wang H, Wei F, Meng S, Cai M, Zhang Y, Xu H, Zhang J, Jin R (2020) Clinical characteristics of coronavirus disease 2019 in Beijing China. PLoS ONE 15(6):e0234764

Carfi A, Bernabei R, Landi F (2020) Persistent symptoms in patients after acute COVID-19. JAMA 324(6):603–605. https://doi.org/10.1001/jama.2020.12603

Castro AKA, Pinheiro PR, Pinheiro MCD, Tamanini I (2011) Towards the applied hybrid model in decision making: a neuropsychological diagnosis of Alzheimer's disease study case. Int J of Comp Intell Syst 4(1):89–99

Delsentoth MP, Falsal AA, Soon C (2020) Mathematics for machine learning. Cambridge University Press, Cambridge

Demsar J, Curk T, Erjavec A (2013) Orange: data mining toolbox in Python. J Mach Learn Res 14:2349–2353

Ferrari F (2020) COVID-19: updated data and its relation to the cardiovascular system. Arq Bras Cardiol 114(5):823

Gupta A, Jain V, Singh A (2021) Stacking ensemble-based intelligence machine learning model for predicting post-COVID-19 complications. N Gener Comput. https://doi.org/10.1007/s00354-021-00144-0

Hilbe JM (2016) Practical guide to logistic regression. Chapman & Hall/CRC, Boca Raton

Hoffman JP (2021) Linear regression models: applications in R. Chapman & Hall/CRC, Boca Raton

Hosmer DW, Lemeshow S (2013) Applied logistic regression. John Wiley & Sons, New York

ICD-10 (2019), https://www.icd.who.int/browse10/2019/en, Accessed 15 Apr 2021

Irwin RS, Lilly CM, Mayo PH, Rippe JM (2018) Intensive care medicine, 8th edn. Wolters Kluwer, Alphen aan den Rijn

Jeon Y, Lin Y (2006) An effective method for high-dimensional log-density anova estimation, with application to nonparametric graphical model building. Stat Sin 16:353–374

Johns Hopkins University—COVID-19 Map—Johns Hopkins Coronavirus Resource Center (jhu.edu)—2021/05/12

Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk A, Eggo RM (2020) Early dynamics of transmission and control of coviD-19: a mathematical modelling study. Lancet Infect Dis 20(5):553–558

Larichev OI, Moshkovich HM (1994) An approach to ordinal classification problems. Int Trans Oper Res 1(3):375–385

Larichev OI, Moshkovich HM (1995) ZAPROS-LM—a method and system for ordering multiattribute alternatives. Eur J Oper Res 82:503–521

Larichev OI, Moshkovich HM (1997) Verbal decision analysis for unstructured problems. Kluwer Academic Press, Dordrecht

Lauer AS, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 172(9):577–582

Liang T (2020) Handbook of Covid-19 prevention and treatment. Zhejiang University School of Medicine, Hangzhou

Machado TCS, Pinheiro PR, Tamanini I (2014) OrclassWeb: a tool based on the classification methodology ORCLASS from verbal decision analysis framework. Math Probl Eng 2014:238168. https://doi.org/10.1155/2014/238168

Mao L, Jin H, Wang M, Hu Y, Chen S, He Q, Chang J, Hong C, Zhou Y, Wang D, Miao X, Li Y (2020) Neurologic manifestations of hospitalised patients with coronavirus disease 2019 in Wuhan China. JAMA Neurol 77(6):683–690

Marinho FP, Loyola IS, Monteiro IOF, Carvalho MGS, Garcia JAD, Silvério ACP, Santos GB (2021) Interrelationship between COVID-19 and diabetes mellitus: a syst review. Res Soc Dev 10(2):e4810212191

Mitchell T (1997) Machine Learning Definition, McGraw-Hill Science/Engineering/Math, http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html

Mitchell T (2011) For pioneering contributions and leadership in the methods and applications of machine learning. National Academy of Engineering

Pant A, Introduction to Logistic Regression. Medium. Towards Data Science. 2019. 685 https://www.towardsdatascience.com/introduction-to-logistic-regression-66248243c148. Accessed Oct 2021

Pecly IMD, Azevedo RB, Muxfeldt ES, Botelho BG, Albuquerque GG, Diniz PHP, Silva R, Rodrigues CIS (2021) COVID-19 and chronic kidney disease: a comprehensive review. Braz J Nephrol 43:383

Perrin R, Riste L, Hann M, Walther (2020) Into the looking glass: post-viral syndrome post-Covid-19. Med Hypotheses 144:110055. https://doi.org/10.1016/j.mehy.2020.110055

Sahami M (1996) Learning Limited Dependence Bayesian Classifiers. In KDD-96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 335–338, Menlo Park, CA: AAAI Press

Santana AV, Fontana AD, Pitta F (2021) Pulmonary rehabilitation after COVID-19. Jornal Bras. de Pneumologia, 47(1)

Scornet E, Biau G, Vert J (2015) Consistency of random florests. Ann Stat 43(4):1716–1741

Shmueli G (2010) To explain or to predict? Stat Sci 25(3):289–310. https://doi.org/10.1214/10-sts330

Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, Felix SEB, Tie Y, Fullerton KE (2020) Coronavirus disease 2019 case surveillance. Morb Mortal Wkly Rep 69(24):759–765

Turban E, Sharda R, Dursun D (2011) Decision support and business intelligence syst. Prentice-Hall, Hoboken

Verma A, Rathi B (2022) Machine learning based predictive model and systems-level network of host-microbe interactions in post-COVID-19 Mucormycosis. Microb Pathog 162(1):105324. https://doi.org/10.1016/j.micpath.2021.105324

Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR (2020) CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. IEEE Access. https://doi.org/10.1109/access.2020.2994762

Wilmott P (2019) Machine learning: an applied mathematics introduction. Panda Ohana Publishing, Oxford

Wolper DH (1997) No free lunch theorems for optimisation. IEEE Trans Evol Comput 1(1):67–82. https://doi.org/10.1109/4235.585893

World Health Organization (2001) International classification of functioning, disability and health. World Health Organization, Geneva, p 302

World Health Organization. Coronavirus disease (COVID-2019) situation reports. 2020. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.